EMPLOYING PSYCHOACOUSTIC MODEL FOR DIGITAL AUDIO WATERMARKING

CHAI SIEW LI

THESIS SUBMITTED IN FULFILMENT OF THE DEGREE OF COMPUTER SCIENCE

FACULTY OF COMPUTER SYSTEM AND SOFTWARE ENGINEERING

2013

ABSTRACT

This thesis discusses about digital audio watermarking by employing psychoacoustic model to make the watermarked signal inaudible to the audience. Due to the digital media data able to distribute easily without losing of data information, thus the intellectual property of musical creators and distributor may affected by this kind of circumstance . To prevent this, we propose the usage of spread spectrum technique and psychoacoustic model for embedding process, zero-forcing equalization and detection and wiener filtering for extracting process. Three samples of audio signal have been chosen for this experiment which are categorized as quiet, moderate, and noise state signal. The findings shows that our watermarking scheme achieved the intended purposes which are to test digital audio watermarking by employing psychoacoustic model, to embed different length of messages to test on accuracy of extracted data and to study the suitability on using hash function for verification of modification attacks.

ABSTRAK

Tesis ini membincangkan tentang audio digital watermarking dengan menggunakan model psychoacoustic untuk membuat isyarat tera air didengar kepada penonton. Oleh kerana data media digital boleh mengedar mudah tanpa kehilangan maklumat data, dengan itu harta intelek pencipta muzik dan pengedar boleh dipengaruhi oleh jenis ini keadaan. Untuk mengelakkan ini, kami mencadangkan penggunaan penyebaran teknik spektrum dan model psychoacoustic untuk proses menerapkan, sifar memaksa penyamaan dan pengesanan dan sosis penapisan untuk proses mengekstrak. Tiga sampel isyarat audio telah dipilih untuk percubaan ini yang dikategorikan sebagai tenang, sederhana, dan isyarat keadaan bunyi. Hasil kajian menunjukkan bahawa skim kami watermarking mencapai tujuan yang dimaksudkan iaitu untuk menguji Mekatronik audio digital dengan menggunakan model psychoacoustic, untuk menanamkan panjang yang berbeza mesej untuk menguji ketepatan data diekstrak dan mengkaji kesesuaian menggunakan fungsi hash untuk pengesahan serangan pengubahsuaian.

TABLE OF CONTENTS

LIST OF TABLES

Table Number **Page**

LIST OF FIGURES

Figure Number **Page**

LIST OF ABBREVIATIONS

ASCII – American Standard Code for Information Interchange
CD – Compact Disc
dB – Decibel
DCT – Discrete Cosine Transform
DFT – Discrete Fourier Transform
DSP – Digital Signal Processing
DWT – Discrete Wavelet Transform
FFT – Fast Fourier Transform
GA – Genetic Algorithm
GUI – Graphical User Interface
HAS – Human Auditory System
Hz – Hertz
IFFT – Inverse Fast Fourier Transform
IT – Information Technology
Kbps – Kilo Byte per Second
kHz – Kilo hertz
LSB – Least Significant Bit
MP3 – MPEG 1 Audio Layer III
SHA – Secure Hash Algorithm
SNR – Signal-to-Noise-Ratio
WAV – Waveform Audio Form

CHAPTER 1 INTRODUCTION

1.1 Introduction

In twentieth century, it is the rise of the digital age and there are many kind of technology had been founded in this digital world. When it comes to digital age, we can say that almost every house, company, restaurant and shop has at least one computer or laptop and everything is digitalized or in digital format. Digital format is a format system that uses binary code which is 0 and 1 only to interpret data received and data to be sent. Other than that, to fight for enrichment, most of the IT expert spends a lot of time in order to make the evolution of technology. Recently, the rapid development of the internet had influence the economy in many aspects such as music production and film production. All the multimedia data in digital format like images, audio, and video can be copied and compressed easily without decrease the fidelity of the data. Hence, with the ease of distribution and duplication, most of the people download illegal copy of the digital media products from the internet.

First of all, there are two famous information hiding techniques which frequently used by the people now, whose are Steganography and Watermarking. Steganography is the art and sciences of writing hiding information in a way that prevents people from detect the hidden messages (Krenn, 2004). Whereas, watermarking is the process of embed a message or a new signal into a host signal. Usually, steganography methods are not robust or with only limited robustness against modification or transmission of the data. On the other hand, watermarking has the additional notion of resilience against attempts to remove the hidden data and other possible attacks. Some of the people may confuse that whether watermark and digital watermark are the same. In fact, when it comes to digital watermark it means that the

watermarked data is in digital format. Digital watermarking is a technique by which copyright information is embedded into the host signal in a way that the embedded information is imperceptible, and robust against intentional and unintentional attacks. (Wang, Niu, Yang, 2009)

By now, the technology of digital audio watermarking will be discovered in this research. Digital Audio Watermarking is a technology to hide information in an audio file without the information being audible to the listener, and without affecting in any way the audio quality of the original file. In addition, to increase the quality of watermarked audio, psychoacoustic model will be employed as psychoacoustic model is a model that designed to take advantage of the masking effect in human hearing which is HAS (Naveen, Jhansi rani, 2010).

1.2 Problem Statement

In the new era of technology world, there are a lot of digital media had been revealed nowadays. There are various types of digital media such as digital video, digital audio, digital images and others. The fast growth of the Internet and the maturity of audio compression techniques enable the promising market of on-line music distribution (Wu, Su, Jay Kuo, 2000). However, with the digital technology today also allows lossless data duplication, illegal copying and distribution would be much easier than before. Consequently, the intellectual property of musical creators and distributor may affected by this kind of circumstance. The necessity of protecting copyrighted audio data has significantly increased. Thus, digital audio watermarking is recommended or promoted to be used for copyright protection and owner authentication.

Copyright is the most concern issue currently as there are a lot of people like to copy the intellectual property illegally and this kind of action is known as piracy. In Malaysia, copyright protection is governed by the Copyright Act 1987 (Perbadanan Harta Intelek Malaysia, 2012). Copyright is the legal protection and exclusive right extended to those who create an original work through sufficient skill and effort for a specific period which depends on the type of works. Apart from that, the protection subsists automatically once the work is published or stored in a material form. In

Malaysia, the works eligible for protection are literacy works, musical works, artistic works, films, sound recordings, broadcasts and derivative works (Perbadanan Harta Intelek Malaysia, 2012). If a work qualified for copyright protection, it gives the owner the right to control the use of his work as well as to prevent it from being copied and distribute to others without owner permission. In term of legal copy is the owner given the right to the person who want to use or copy his work and that person only allowed to use it personally and is not allowed to use it for trading or others thing that is related to profit. By using digital watermarking method, we are able to trace and tackle the people who make the illegal copy no matter it is make a copy into CD or distribute to the internet. Once we extract information from the illegal copy which is watermarked copy, we are able to prove whether it is an authorized copy.

Moreover, owner authentication is uses a set of data or information that can be used to identify or prove the ownership. There will be some problems may arise when people are able to change the set of data or information easily in order to get the ownership. In digital watermarking, the embedded information may not be easily removed or changed as it required the secret key and it is robust to attacks.

1.3 Objectives

Below show the objectives of this research:

i. To test digital audio watermarking by employing psychoacoustic model
ii. To embed different length of messages to test on accuracy of extracted data
iii. To study the suitability on using hash function for verification of modification attacks

1.4 Scope of Study

Below show the scope of this research:

i. Spread spectrum technique and psychoacoustic model will be used as watermark embedding scheme
ii. Zero-forcing equalization and detection, and wiener filtering method will be used as watermark extracting scheme
iv. Hash sum function will be used to hash the watermark message

In this project, psychoacoustic modeling will be used and integrated with the Spread Spectrum technique. The purpose for this integration is to enhance the Spread Spectrum technique which is generated an audible noise and presence of strong distortion for the watermarked audio. In addition, by using spread spectrum method it will increase the robustness for watermark as the presence of pseudo-random generator. In addition, zero-forcing equalization and detection will be used for watermark extracting process and wiener filtering as an enhancement method that support the zero-forcing equalization and detection method. To study the suitability on using hash function for verification of modification attacks, hash sum function will be used to hash the watermark message and embed it into the audio signal and extract it from the watermarked signal. By using those techniques, watermark can be added and detected easily. I am here to study about the transparency for the audio watermarking and to investigate the audio quality between original audio and watermarked audio. Overall, in this digital audio watermarking system, users are able to watermark or detect watermarked audio. There is no boundary in user and place. It can be used by anyone at every place.

1.5 Thesis Organization

This thesis consist total of six (6) chapters. Chapter 1 will introduces the project which includes problem of background, objectives and scope of project. It will introduce digital audio watermarking, the background issues, and overview of the project. Chapter 2 will discuss on existing research or system that done by other researchers and explain which techniques or methods more suitable to be used in this project. Chapter 3 will discuss the overall approach and model flow of research which includes the method, technique or approach to be used in this project. It will explain the psychoacoustic model, spread spectrum technique, and ASCII encoding and decoding algorithm. Chapter 4 will discuss on development for the model flow of research which includes data collection, process, and analysis. Chapter 5 will discuss about the results after implement the proposed model flow for the research and discuss on the findings includes constraints and limitation for the proposed watermarking scheme. Chapter 6 will conclude the whole research that has been done and future direction.

CHAPTER 2 LITERATURE REVIEW

2.1 Watermarking

Mohanty, S.P. (1999) had summarized types of watermarking which shown in Figure 2.1.

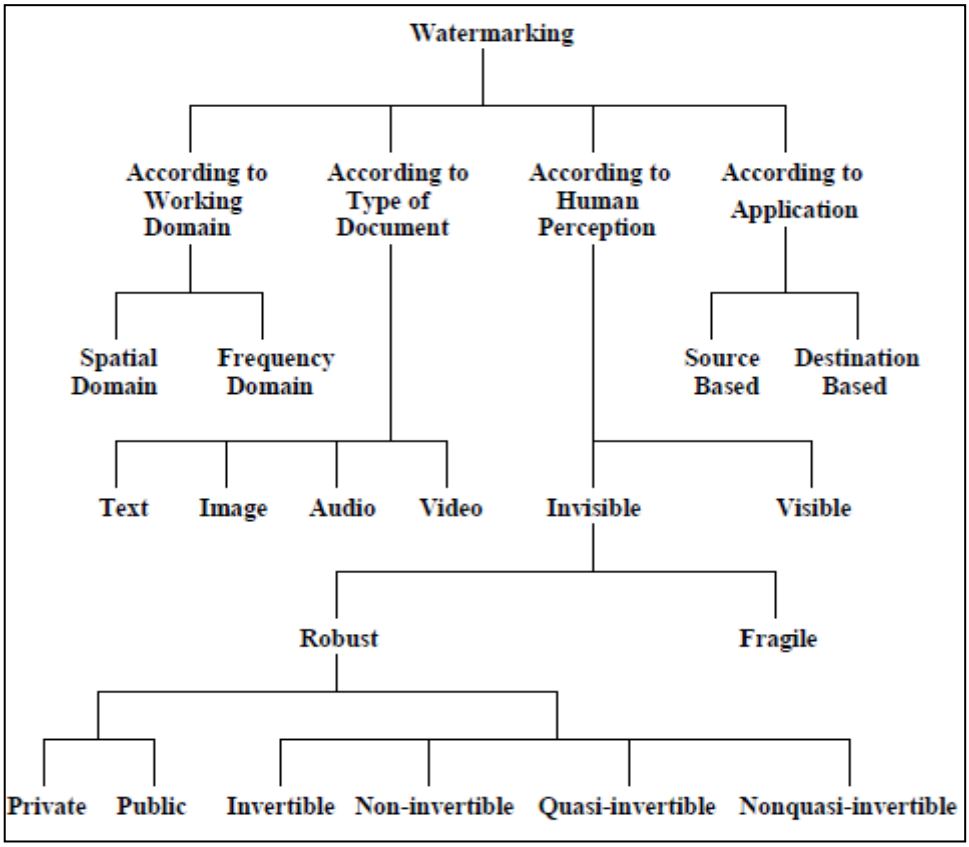

Figure 2.1 :Types of watermarking techniques

Watermarking can be applied to type of document like text, image, audio and video. Geetha and Vanitha Muthu (2010) had proposed Embedding Text in Audio

Signal (ETAS) model which is a very basic description of the audio steganography process in the sender side and receiver side. They use the LSB coding method to encode the message in audio signal. Figure 2.2 shows the ETAS model. Furthermore, table 2.1 shows the evaluation of Steganography requirements associated with cover medium that had been founded by them. Other than that, they also make a comparison between cover medium of text merged with in terms of size which shown in Figure 2.3. Anyway, they stated that ETAS model is able to ensure secrecy with less complexity at the cost of same memory space as that of encrypted text and the user is able to enjoy the benefits of cryptography and steganography combined together without any additional overhead.

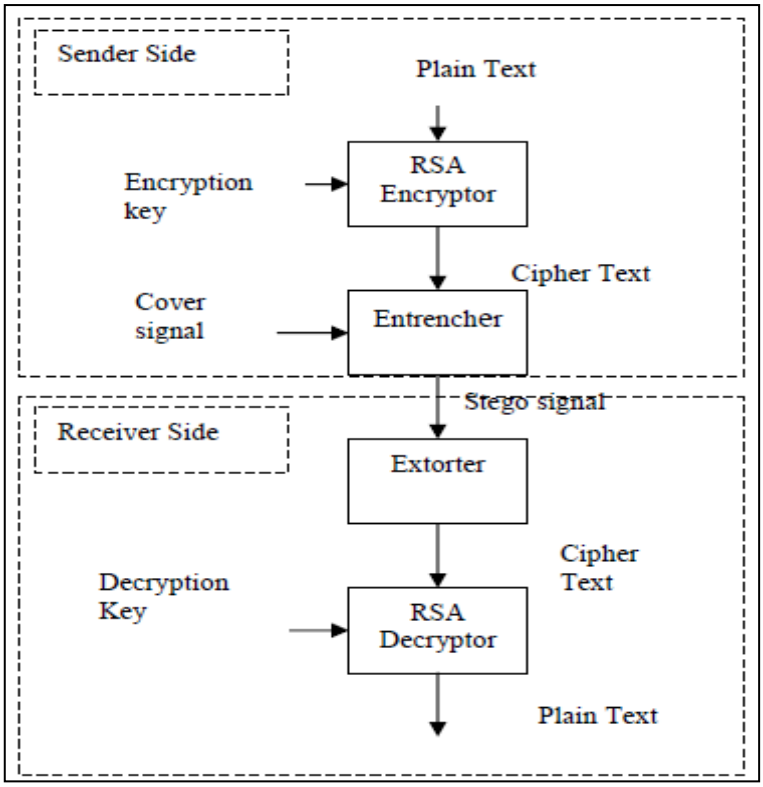

Figure 2.2 :ETAS Model

Table 2.1 :Evaluation of Steganography requirements associated with cover medium (Geetha and Vanitha Muthu, 2010)

	Plain Text	Image	Audio	Video
Invisibility	Medium	High	High	High
Payload Capacity	Low	Low	High	High
Robustness against Statistical Attacks	Low	Medium	High	High

Robustness against Text Manipulation	Low	Medium	High	High
Variation in file size	Medium	Medium	High	Medium

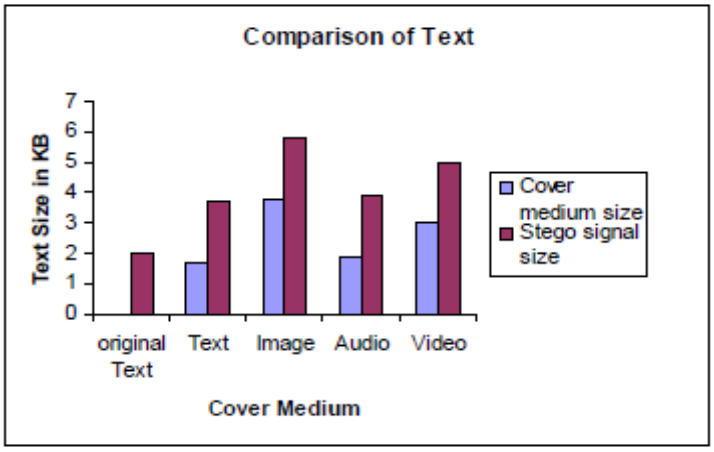

Figure 2.3 :Comparison of text merged with in terms of size (Geetha and Vanitha Muthu, 2010)

Based on the result on above, we can know that watermarking in audio is the better to embed the secret information. Hence, in this research the audio document type will be chosen for undergo watermarking process. Other than that, Agbaje, Akinwale and Njah (2011) also stated that audio signals are represented by much less samples per time interval, which indicates that the amount of information capacity that can be embedded robustly and inaudibly in audio files is much lower than the amount of information that can be embedded in visual files.

2.1.1 Cryptography versus Steganography

Amin *et al.* (2003) stated that the purpose of cryptography and steganography is to provide secret communication. However, steganography is not the same as cryptography. Cryptography hides the contents of a secret message from malicious people, whereas steganography even conceals the existence of the message. Basically, cryptography offers the ability of transmitting information between persons in a way that prevents a third party from reading it. In cryptography, the structure of a message is scrambled to make it meaningless and unintelligible unless the decryption key is

available. Whereas in steganography, it does not alter the structure of the secret message but hide it inside a cover-image so that it cannot be seen. In other words, cryptography is an encrypting process while steganography is a hiding process.

2.1.2 Steganography versus Digital Watermarking

Mohanty, S.P. (1999) said that steganography and digital watermarking primarily differ by intent of uses. A watermark can be perceived as an attribute of the carrier (cover). It may contain information such as copyright, license, tracking and authorship. While in case of steganography, the embedded message may have nothing to do with the cover. In steganography an issue of concern is bandwidth for the hidden message whereas robustness is of more concern with watermarking.

2.2 Working Domain

There are two type of working domain which is spatial and frequency domain. According to Singh (2011), spatial domain is manipulating or changing an image representing an object in space which uses statistical properties of each pixel and its immediate surrounding pixels in the host image, and also the statistical properties of the host image and that of the image to be embedded (watermark), as the pixels in the host image are replaced one by one by the pixels in the watermark image. One of the spatial domain techniques is LSB in which message is embedded in the least significant bit. While in transform domain, watermark is embedded in frequency domain of a signal such as DCT, DFT, DWT domain coefficients. Transform domain methods hide messages in significant areas of the host image which makes them more robust to attacks.

According to Cai and Chen (2011), the time domain masking effect refers to the weak voice fore-and-aft a stronger voice cannot be detected by the human's ears, that is, they would be "masking" out. The frequency domain masking effect, also known as the same time masking, indicating when two signals of nearly close frequency are concurrently working, the weak sound will be masked by strong sound and becomes unaware, and then the masking sound has some effect during the working of masking

effect, which is a much stronger masking effect. To ensure the efficient extraction of the watermarking, it can be reduce the energy of an embedded watermark as much as possible, which make it fully "submerged" in the carrier (cover) data.

Anyway, most of the audio watermarking is working in transform domain. One of the famous transform methods being used in audio watermarking is FFT. The Fourier Transform is a mathematical operation which converts time-domain signals to the frequency domain. Being able to visualize a signal in the frequency domain offers many benefits over time-domain representations. Frequency domain representations allow individual frequency components contained within a signal to be viewed including modulation sidebands, distortion effects and spurious frequency components. For example, figure 2.4 show a square wave is composed of a fundamental frequency and all it odd harmonics. The level for each harmonic decreased in amplitude with harmonic number. Viewed in the time domain representation, the square wave is gives no indication of its composition. Viewed in the frequency domain all frequencies components are displayed along with their relative amplitude (Aeroflex, 2012). Furthermore, the DFT is a version of the Fourier Transform which can be applied to sampled time-domain signals. The DFT produces a discrete frequency spectrum which is an amplitude levels at discrete frequencies. Whereas, the FFT is a development of the DFT which removes duplicated terms in the mathematical algorithm to reduce the number of mathematical operations performed. In this way, it is possible to use large numbers of samples without comprising the speed of the transformation. In other words, FFT is a fast version of DFT it transform speed is faster than the DFT.

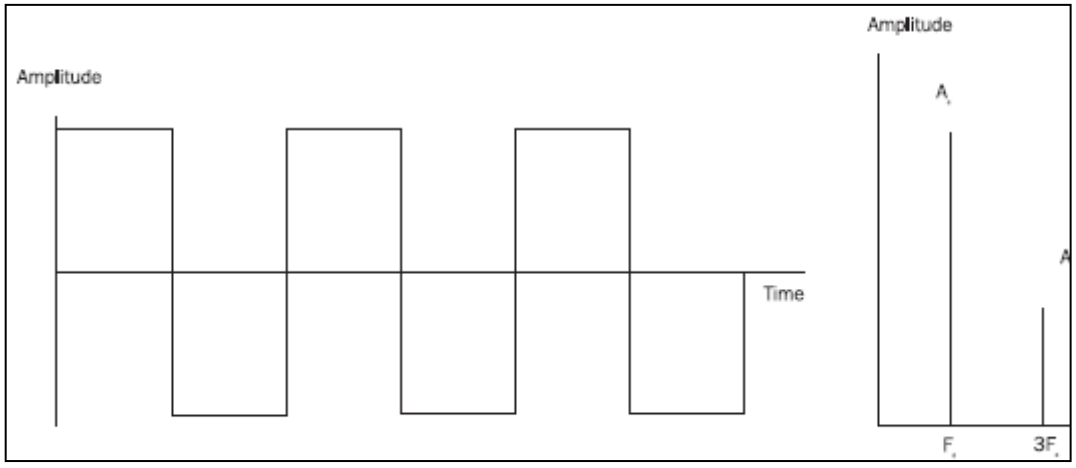

Figure 2.4 :Time and frequency domain representations of a square-wave signal

2.3 Watermarking Technique Requirements

According to Singh (2011), there are four watermarking technique requirements which are robustness, non-perceptibility, verifiability and security.

- Robustness

  Robustness refers to that the watermark embedded in data has the ability of surviving after a variety of processing operations and attacks. Then, the watermark must be robust for general signal processing operation, geometric transformation and malicious attack. The watermark for copyright protection does need strongest robustness and can resist malicious attacks, while fragile watermarking; annotation watermarking do not need resist malicious attacks.

- Non-perceptibility

  Watermark cannot be seen by human eye or not be heard by human ear, only be detected through special processing or dedicated circuits.

- Verifiability

  Watermark should be able to provide full and reliable evidence for the ownership of copyright-protected information products. It can be used to determine whether the object is to be protected and monitor the spread of the data being protected, identify the authenticity, and control illegal copying.

- Security

  Watermark information owns the unique correct sign to identify, only the authorized users can legally detect, extract and even modify the watermark, and thus be able to achieve the purpose of copyright protection.

Watermark information owns the unique correct sign to identify, only the authorized users can legally detect, extract and even modify the watermark, and thus be able to achieve the purpose of copyright protection.

2.4 Characteristic of Audio Watermarking Techniques

Audio watermarking techniques have three characteristics which are inaudibility, robustness and bit rate. Figure 2.4 shows the magic triangle which represent those characteristics (Cvejic, N.,2004). Inaudibility is presented in the upper portion of the triangle as it is the top requirement of audio watermarking process. While, the other two requirements cannot achieve together so it presented in two corners of the triangle respectively. For example, if the data rate is high then the robustness is low while if the robustness is high then the data rate is low.

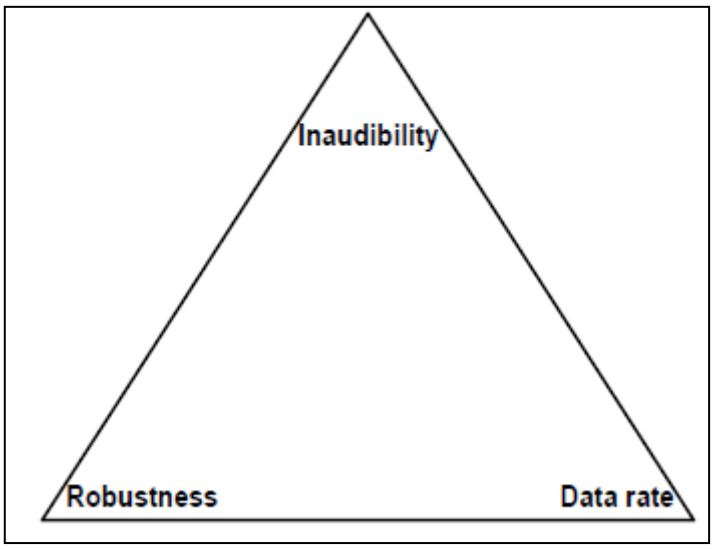

Figure 2.5 :Magic Triangle (Cvejic, N., 2004)

2.5 Audio Watermarking Technique

Kim, Choi, Seok and Hong (2004) state that there are four categories scheme for audio watermarking those are spread-spectrum scheme, two-set scheme, replica scheme and self-marking. Spread-spectrum embeds pseudo-random sequence and detects by calculating auto-correlation. While two-set scheme which is also known as patchwork scheme exploits the differences between two or more sets. Replica scheme uses the replica of the original audio clips both in embedding and detection phases. Echo hiding is a good example of replica scheme. Last one is self-marking scheme which can be used especially for synchronization or for robust watermarking, for example, against time-scale modification attack. Time-scale modification refers to the process of either compressing or expanding the time-scale of audio. The basic idea of the time-scale

modification watermarking is to change the time-scale between two extrema (successive maximum and minimum pair) of the audio signal. Such four seminal works have improved watermarking schemes remarkably. Pseudo-random sequence has statistical properties similar to those of a truly random signal, but it can be exactly regenerated with knowledge of privileged information. Good pseudo-random sequence has a good correlation property such that any two different sequences are almost mutually orthogonal. Thus, cross-correlation value between them is very low, while auto-correlation value is moderately large. In addition, Table 2.2 shows the comparison of various watermarking technique written by Agbaje, Akinwale and Njah (2011).

Table 2.2 :Comparison of various watermarking techniques (Agbaje, Akinwale and Njah, 2011)

Techniques	Advantages	Disadvantages
Spread spectrum	Easy to implement	It requires time consuming psychoacoustic shaping to reduce audible noise, susceptible to time-scale modification attacks and difficulty in synchronization
Quantization	Easy to implement and robust against noise to a particular threshold	Not robust against attacks
Two set	-	-
Replica method	Immunity to synchronization attacks	-
Echo hiding	imperceptibility	High complexity due to ceptrum or autoceptrum computation during detection and echo can be detected without prior knowledge

Moreover, according to Kiah et al. (2011), there are number of ways to hide the information or data into audio such as phase coding, spread spectrum, echo data hiding, patchwork coding, low-bit encoding and noise gate. Table 2.3 shows the advantages and disadvantages of the techniques that written by Kiah et al. (2011).

Table 2.3: Advantages and Disadvantage of the techniques

Approach	**Summary**	**Advantage and Disadvantage**
Low–bit Encoding	Low-bit encoding considered as the earliest techniques implemented in the information hiding of digital audio. It is the simplest technique to embed data into other data structures such as data of audio in image file or data of image in audio file. Low-bit encoding, can be done by replacing the LSB of each sampling point by a coded binary string (hidden data)	The major advantage of Low-bit encoding are: 1. High watermark channel bit rate 2. Low computational complexity of the algorithm compared with others techniques 3. No computationally demanding transformation of the host signal, therefore, it has very little algorithmic delay The major disadvantage is that the method are: 1. Low robustness, due to the fact that the random changes of the LSB destroy the coded watermark 2. it is unlikely that embedded watermark would survive digital to analogue and subsequent analogue to digital conversion
Phase Coding	Phase Coding watermarking works by substituting the phase of an initial audio segment	The major advantage of Phase Coding are: 1. Basic technique

	with a reference phase, this phase represents the hidden data. The phase of subsequent segments is adjusted in order to preserve the relative phase between segments	The major disadvantage is that the method are: 1. Phase coding method is a low payload because the watermark embedding can be only done on the first block. 2. The watermark is not dispersed over the entire data set available, but is implicitly localized and can thus be removed easily by the attackers
Spread Spectrum Technique	Spread spectrum (SS) is technique designed to encode any stream of information via spreading the encoded data across as much of the frequency spectrum as possible even though, there is interference on some frequencies, SS allows the signal reception.	The major advantage of Spread Spectrum are: 1. Difficult to detect and/or remove a signal 2. Provide a considerable level of robustness The major disadvantage is that the Spread spectrum are: 1. Spread spectrum technique used transform functions (e.g. DFT, DCT, or DWT) with appropriated inverse transform function, which can cause a delay. 2. Spread spectrum is not a visible solution for real time applications
Patchwork Coding	Patchwork Coding considered as one of the earliest generation for digital	The major advantage of Patchwork Coding are: 1. Patchwork based watermarking

	watermarking schemes. Patchwork Coding can be done via embedding the watermark in the audio using time domain or frequency domain. In the literature, several approaches of Patchwork Coding have been proposed on frequency domain using linear transformations, such as DWT, DFT and DCT. Frequency or time domain watermarking schemes directly tinker with sample amplitude of audio to embed the watermark	scheme has been confirmed as an valuable to those common signal processing operations, such as low-pass filtering, image/audio compression, and so on. The major disadvantage is that the Patchwork are: 1. An attack called "curve-fitting attack" has been successfully implemented for patchwork watermarking scheme. 2. Patchwork watermarking scheme is sensitive to various synchronization attacks
Echo technique	Echo technique embeds data into a host audio signal by introducing an echo; the hidden data can be adjusted by the two parameters: amplitude and offset, the two parameters represent the magnitude on time delay for the embedded echo, respectively. The embedding process uses two echoes with different offsets, one to represent the binary datum "One" and the other to represent the binary datum "Zero".	The major advantage of Echo are: 1. The main advantage of echo hiding is that the echo detection technique is easy to implement. The major disadvantage is that the echo hiding technique are: 1. More complicated computation is required for echo detection. 2. Echo hiding is also prone to inevitable mistakes, such as the echo from the host signal itself may be treated as the embedded echo. 3. If the echo added has smaller amplitude, then the cepstrum peak would be covered by the